Self-Reinforcing Gating via Directional Alignment

Consider a feedforward neural network composed of linear transformations followed by gating nonlinearities (e.g., ReLU or multi-branch gating). Suppose inputs are drawn from a distribution with recurring directional structure, and parameters are updated by gradient-based learning.

Then:

1. **Directional Selectivity**
   Each linear unit acts as a directional filter: its activation magnitude increases with alignment between the input vector and its weight vector.
2. **Induced Gating Partition**
   The network induces a partition of input space into regions indexed by gating patterns (activation states), where each region corresponds to a distinct effective linear mapping.
3. **Reinforcement of Activation Patterns**
   During training, inputs that initially align with a given set of weight directions are more likely to activate the same gating pattern. Gradient updates then strengthen those weight directions, increasing future alignment for similar inputs.
   Hence, activation patterns become **self-reinforcing**.
4. **Stabilization of Routing**
   As alignment magnitudes grow, activations move away from decision boundaries, making gating patterns locally stable under small input perturbations.
   Therefore, similar inputs are routed consistently through the same sequence of linear transformations.
5. **Specialization of Linear Mappings**
   Each stable gating pattern is repeatedly exposed to a coherent subset of inputs sharing directional similarity. The associated effective linear mapping adapts to this subset, yielding **context-specific specialization**.
6. **Implicit Weight Sharing Constraint**
   Because different gating patterns share underlying parameters, updates arising from one region constrain others. Training therefore favors linear mappings that are **mutually compatible across neighboring regions**, rather than arbitrary independent associations.
7. **Generalization via Directional Clustering**
   Inputs that differ in irrelevant components but share alignment with dominant weight directions are mapped to the same or nearby gating patterns.
   Consequently, the network generalizes by grouping inputs according to **directional similarity** rather than exact identity.

Corollary (Emergent Structured Hashing Interpretation)

Under the same conditions, the network implements a **learned, locality-sensitive mapping** from inputs to linear transformations, where:

- Gating patterns function as **hash keys**,
- Effective linear mappings function as **values**, and
- Training enforces **consistency across neighboring keys** through shared parameters.

Informal Summary

Gradient learning in gated linear networks amplifies initially preferred directions, causing stable activation patterns to emerge. These patterns define consistent routing paths, each associated with a specialized but shared linear transformation, yielding a structured and generalizing form of input-dependent computation.